RNA-mediated Double-Stranded Break repair in mammalian cells

A thesis
Presented to
The Academic Faculty

By

Matthew B Taylor

In Fulfillment of the requirements
for the Honors Thesis Option
for a Bachelor's Degree in the
School of Biology

Georgia Institute of Technology
April 2011

RNA-mediated Double-Stranded Break repair in mammalian cells

Approved by:

Dr. Francesca Storici, Advisor
School of Biology
Georgia Institute of Technology

Dr. Gregory Gibson, Second Reader
School of Biology
Georgia Institute of Technology

Date Approved:

Table of Contents

Table of Figures

Glossary

DSB: Double-stranded break. DNA damage event involving the cleavage of bonds in the backbone of both strands of a DNA molecule

Topocloning: Method by which a topoisomerase is used to add PCR product into a plasmid, which can then be replicated within bacterial cells. This allows researchers to study an individual DNA molecule replicated by PCR, as opposed to the aggregate of a PCR reaction.

PCR: Polymerase Chain Reaction. This method utilizes replicative enzymes to multiply copies of a tract of DNA.

Plasmid: A circular molecule of DNA. Plasmids can be taken up by cells and are frequently used as vectors for passing genes to cells being studied.

Oligonucleotide: A Synthetic strand of DNA and/or RNA.

Abstract

This study seeks to examine the capability of RNA to introduce mutations into genomic DNA, and to provide insight into the enzyme(s) responsible for this phenomenon. To do this, a double-stranded break is induced in HEK 658D cells, and oligonucleotides containing RNA tracts with homology to the area around the break are introduced to examine if they can mediate repair of this area. In the first part of the study, oligonucleotides are introduced that contain a single-nucleotide substitution at the site of the break, and if the RNA is used as a template for repair, then this nucleotide should be incorporated into the genome of the cell. Oligonucleotides are introduced that contain either all-DNA, or DNA with a 6-nucleotide RNA tract at the break. These two classes of oligonucleotides are homologous to either the forward or reverse strands of the area around the break. This allows us to examine if there is a strand bias for repair, and if that bias is the same for DNA- and RNA-mediated repair, which will provide clues into the enzyme involved in repair.

Chapter 1
Introduction

DNA (deoxyribonucleic acid) is a double-stranded molecule consisting of two strings of molecules called nucleotides. Each nucleotide is capable of binding to one other nucleotide (there are 4 total), and in a complete double-stranded DNA molecule each strand is composed of nucleotides that bind to the nucleotides of the opposite strand. These strands are referred to as the forward and reverse strands, where the forward strand is used as a template for gene expression, and the reverse strand is an "antisense" copy of that sequence. The sequences of these two strands are referred to as "complimentary", since each nucleotide lines up with its match when the strands are bound together.

The structure of DNA gives it its vital function: the ability to "code" for molecules that perform tasks in every living cell. The exact order of nucleotides on a strand of DNA carries a message that can be used to direct the synthesis of molecules of RNA (ribonucleic acid). RNA is similar to DNA from a chemical standpoint, and one of its functions in the cell is as a template for protein synthesis. The sequence of RNA can be used as a set of instructions to create proteins. RNA was once thought of as a messenger only, but this is not the case. RNA molecules are involved in enzymatic activities, such as the production of Proteins (Ribosomes and tRNA) and in gene silencing (siRNA/RNAi), as well as several other functions that are just being elucidated. Discovery of these vital functions has led to a surge in research to learn about the properties of RNA and how it can interact with other elements inside a cell. One of these newly discovered functions of RNA, gene silencing through RNAi, is currently being explored in a staggering array of studies. The ability to block the expression of a gene is extremely useful in many types of

genetic analysis, as it gives a scientist the ability to all but turn off genes at will without the arduous process of engineering organisms missing that gene.

Past work by Storici et. al. exploiting the use of RNA-containing oligonucleotides has shown that RNA is capable of mediating the repair of a DNA damage event known as a double-stranded break (DSB). During these events, a break occurs in between nucleotides on both strands of a DNA molecule, similar to cutting a braided rope with shears. These events can lead to loss or rearrangements of genes, and are frequently fatal to the cell. If an RNA molecule complementary to the area on either side of a DSB is introduced to cells suffering from such a break, the RNA can act as a "molecular band-aid" and aid in the proper repair of the original DNA molecule. Enzymes inside the cell's nucleus are able to use the RNA'S sequence as a template for repair.

This study seeks to examine whether RNA can be used as a template for DNA synthesis during repair of a double-stranded break event in human cells. To this point it is unclear whether RNA has actually been the template for repair, or if perhaps a cDNA intermediate is involved due to the action of a reverse transcriptase. RNA-containing DNA oligonucleotides complementary to a region of the GFP gene will be introduced to cells suffering from a break in that region. These oligonucleotides have been designed with a single-nucleotide difference from the original DNA sequence of the break region, and after the break this inserted nucleotide will fall in between the area of the break. If the RNA in this region is used as a template for repair of the DSB, then this single-nucleotide difference will be incorporated into the repaired GFP gene. DNA molecules used as a template for repair are most effective when they are complimentary to the forward strand of the damaged DNA. In this study, we will compare the repair frequency of RNA molecules complimentary to either the forward or reverse strand with those of DNA molecules

complimentary to either strand. If the strand bias is the same, it will suggest that a similar mechanism works for DNA and RNA, and rule out the possibility that the RNA is reverse-transcribed into a cDNA that acts in the repair.

In a follow-up study, this process will be repeated but with an oligonucleotide that contains a "mismatched" oligonucleotide, a single-nucleotide difference that will be non-homologous to one side of the broken DNA molecule. If this oligonucleotide is used as a template for repair of the DSB, then it is possible that the single-nucleotide change will be incorporated into the GFP gene, showing that RNA can in fact affect the sequence of a DNA molecule.

Chapter 2

Literature Review

Questions about which molecule(s) in a cell held its genetic material virtually ended when Hershey and Chase showed that DNA in viruses, as opposed to proteins, was the purveyor of genetic information (Hershey and Chase 1952). Since the Nobel-prize winning discovery of the structure of DNA by Watson and Crick (1953), scientists have been working on elucidating the properties and interactions DNA takes part in and how its structure contributes to these interactions. By the 1970s, scientists felt they had a firm enough understanding of the process of gene expression to term their understanding the "central dogma of biology" (Crick 1970), which stated the following: DNA was used as a template to transcribe RNA, which was then translated into a protein. The belief at the time was that this flow of information (from gene to product) was unidirectional, and could not be reversed. This misconception did not stand for long, however, as the discovery of reverse transcriptase proved that this flow could be reversed. Reverse transcriptase is most famous for its role in the HIV virus, in which it uses a strand of RNA as a template to produce DNA containing viral genes. It has even been shown that ribonucleotides are incorporated into the genome of most organisms during replication, and the effect of this incorporation on genome stability is still unclear.

Further studies have shown the ability of RNA to interact with DNA directly. First detected in 1990, the 2006 Nobel Prize was awarded for the characterization of small, double-stranded RNA molecules with the ability to bind to homologous segments of mRNA (Fire 2007). These molecules are called siRNA, small interfering RNA, for their ability to silence genes by binding to and partially blocking transcription, and thus reducing expression. This revolutionary discovery

has lead the way into a new interest in studying RNA, a class of molecules once thought to be nothing more than a cellular messenger. This discovery has had immediate medical applications (Caplen and Mousses 2003; Karagiannis and El-Osta 2004), as it has allowed researchers to design RNA molecules to target and silence (turn off) detrimental genes (responsible for genetic diseases).

These interfering molecules work on a very interesting mechanism. Double-stranded RNA molecules that contain a region homologous to mRNA for the desired gene are introduced to cells. The RNA molecule is cleaved by enzymes in the cells and incorporated into a protein complex called the RNAi induced silencing complex (RISC). This complex then binds to free mRNA homologous to the bound RNA and targets those molecules for degradation (Karagiannis and El-Osta 2004).

In 2007, Storici showed that RNA can be used in another purpose: to aid in the repair of DSBs (Storici et al. 2007). When a DNA molecule experiences a DSB, the results can be fatal, if the break is not repaired promptly and precisely. Newly free ends of DNA can be bound randomly to other ends of DNA molecules, causing mutations and genomic rearrangements (Chayot et al., 2010).

The first step of this study is to examine whether RNA can serve as a template for DNA is synthesis during DSB repair. An oligonucleotide containing a 6-nucleotide region of RNA is introduced to a cell suffering from a DSB in the green fluorescent protein (GFP) gene. The RNA-DNA hybrid is homologous to the area around the break, and if it binds as expected, it will provide a template for repair of the break. The RNA segment of this oligonucleotide has a single-nucleotide difference from the strand it binds to. When the oligonucleotide binds to the

chromosomal DNA, the mismatched nucleotide will be directly over a gap in the broken region. We hypothesize that the repair enzymes will synthesize DNA complementary to the RNA "band-aid" in the repaired DNA (figure 1). The cells in which repair occurs will fluoresce green when exposed to a certain wavelength of light, clearly showing whether or not the repair was successful (figure 2). The single-nucleotide mutation introduced by the RNA tract of the oligonucleotide will cause the DNA sequence to code for a *Bss*HII restriction enzyme cut site, which can be easily visualized with electrophoresis.

In addition to the RNA-containing oligonucleotide, this process was also tested with an all-DNA oligonucleotide containing the same nucleotide sequence, and each will be performed with oligonucleotides that are complementary to either the forward or reverse strand of the GFP gene. Preliminary work by our lab has shown that forward-complimentary oligonucleotides can efficiently repair the DSB in GFP and introduce the desired mutation (Shen et. al., 2010) . However, it has not been shown whether the repair synthesis is mediated by a DNA polymerase or a reverse transcriptase. If the DNA and RNA repair pathways show similar bias for the forward as opposed to the reverse oligonucleotide, it will suggest that the RNA containing oligonucleotides are indeed acting as a template in this repair pathway. Conversely, if the RNA-containing oligonucleotide is reverse transcribed into DNA, we expect a reverse bias in repair efficiency by the RNA-containing oligonucleotides, or no bias if a double-stranded cDNA molecule is made. A similar study has been performed on yeast (Storici, et al., 2007), and we expect similar results.

Further studies can explore what enzymes specifically are involved in this repair pathway. Expression of genes coding for polymerases or reverse transcriptases could be knocked down

and then the procedure above repeated to see if loss of polymerases inhibits repair by the RNA-containing oligonucleotides. Another direction would be to see if endogenous (already inside the cell) RNA can transmit information. To do this, a short gene would be inserted into cells coding for RNA that is the same in sequence as the oligonucleotides used above. If a DSB is induced, the RNA produced by the inserted gene could repair the DNA and generate a *Bss*HII site as above. Understanding more about how RNA is able to mediate DNA repair and transmit genetic information to the genome of cells will add a new piece to the genetic toolbox available to scientists, and can hopefully provide new routes to gene therapy.

Chapter 3

Methods

Cells:

All work in this experiment was performed in Human Embryonic Kidney (HEK) 658D cells. The HEK cell line is a human cell line commonly used in molecular biology studies. The 658D variant has a copy of GFP inserted into its genome that is non-functional because it has been interrupted by the insertion of a 35 nucleotide region. This insertion prevents the gene from being functional. Within the 35-base pair region is a site that can be recognized and cut by an enzyme called I-*Sce*I. This enzyme is part of a class of enzymes called homing endonucleases, which are capable of shearing DNA to create a double-stranded break. Cells are incubated at $37^{o}C$

Oligonucleotides:

Oligonucleotides are produced complimentary to either side of an I-*Sce*I site in the middle of the GFP gene. These oligonucleotides are 70 bases long, with 32 bases complimentary to each side of the I-*Sce*I site and a 6 base tract in the middle that codes for a *Bss*HII cut site. There are 4 groups of oligonucleotides:

1. Oligonucleotide complementary to forward strand of area around DSB

   GFP66.F

   5’-

   ATCTTCTTCAAGGACGACGGCAACTACAAGAC**G**CGCGCCGAGGTGAAGTTCGAGGGCGAC

   ACCCTG

2. Oligonucleotide complementary to reverse strand of area around DSB

   GFP66.R

   5’-

   CAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGCGTCTTGTAGTTGCCGTCGTCCTTGAAG

   AAGAT

3. Oligonucleotide complementary to forward strand area around DSB, 6-nucleotide tract of RNA at break

   GR66_6R.F

   5’-

   ATCTTCTTCAAGGACGACGGCAACTACAAGACGCGCGCCGAGGTGAAGTTCGAGGGCGAC

   ACCCTG

4. Oligonucleotide complimentary to reverse strand of area around DSB, 6-nucleotide tract of RNA at break

   GR66.6R.R

5’-

CAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGCGTCTTGTAGTTGCCGTCGTCCTTGAAG

AAGAT

5. Oligonucleotide complimentary to reverse strand of area around DSB, 6-nucleotide tract of RNA at break, single-nucleotide mismatch 15 nucleotides from area of break

GR66_MM1D.F

5’-

ATCTTCTTCAAGGACGACGGCAACTACAAGACGCGCGCCGAGGTGAAATTCGAGGGCGAC

ACCCTG

Transfection:

Cells are "seeded" at a low density (~50,000 per well on a 24-well plate) and allowed to grow overnight. A mixture of media, polyethelenimine, and DNA/RNA are added evenly to each well. Once green cells can be observed in the experimental groups (usually a few days), the number of green vs. non-green cells is counted. Green cells are sorted and their genomic DNA is extracted. This is run in PCR to amplify the GFP gene and then digested with *Bss*HII and I-*Sce*I to determine if the repair and modification occurred as hypothesized.

Statistics:

Cells are passed through a flow cytometer, which counts incidences of fluorescence.

Approximately 100,000 cells are counted to determine the frequency of repair.

Topocloning:

PCR product from amplification of the repaired region is ligated into a plasmid vector (as per Zero-blunt topocloning kit). E. coli are transformed with this plasmid vector and grown on selective media following Invitrogen kit protocol. Single-cell colonies are isolated and grown in selective LB broth media overnight. Plasmid from each of these cultures is isolated using Midi-prep kit (Quiagen), and then digested with B*ss*HII and I-*Sce*I to determine if the repair and modification occurred as hypothesized. Colonies showing expected digestion patterns are PCR amplified and sent in for sequencing.

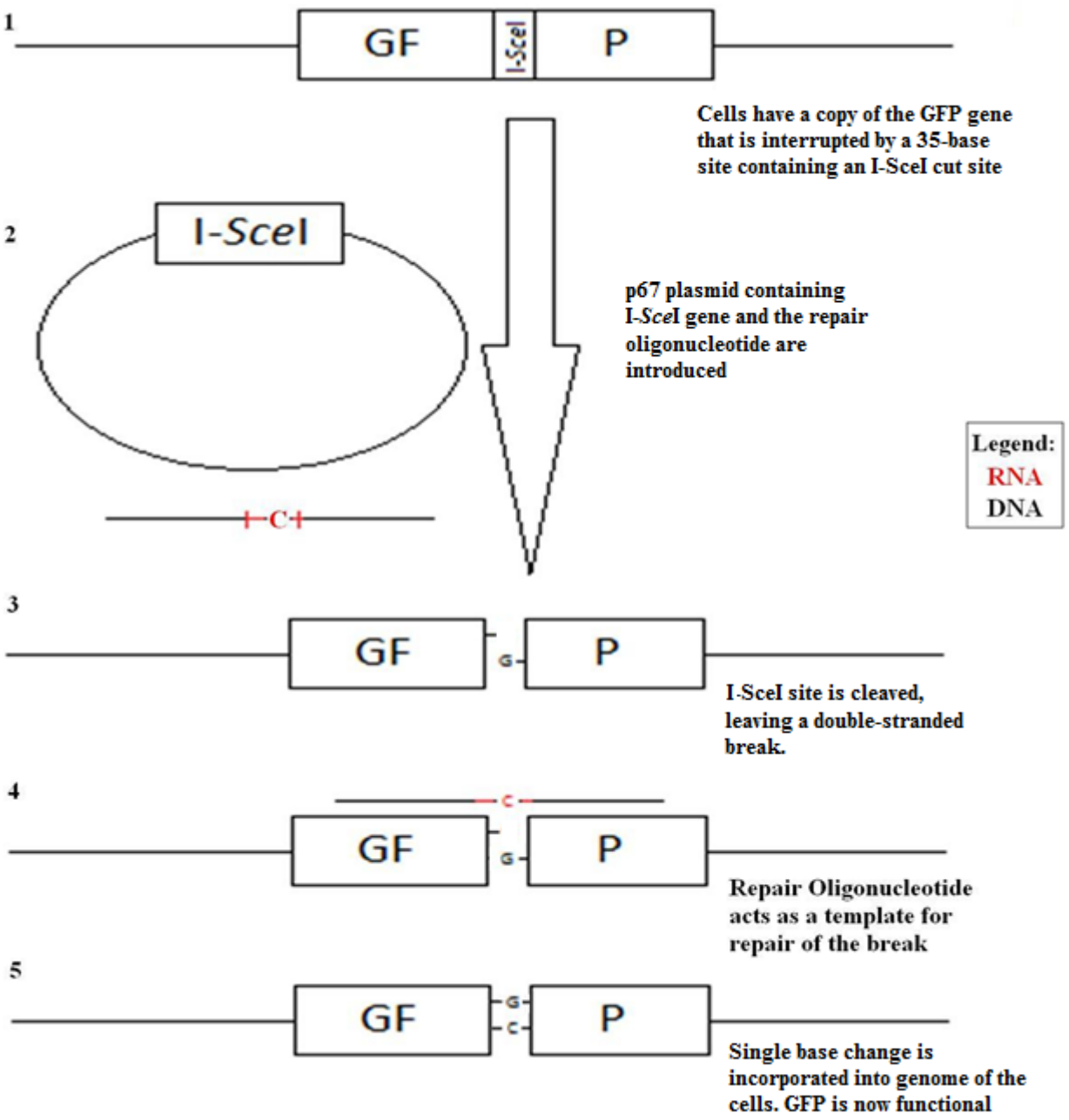

Figure 1—Repair of DSB by oligonucleotide. I-*Sce*I site is cleaved by I-*Sce*I enzyme, leaving two free ends in the GFP. Oligonucleotide hybridizes to the area around the break, holding the two ends in place for repair enzymes to fix the break. 6-base pair region on the oligonucleotide at the break region is composed of RNA and contains a single-base mutation. If the repair pathway for the DSB utilizes the oligonucleotide as a template, the newly-repaired gene will have a single-base mutation that introduces a *Bss*HII cut site.

Chapter 4

Results

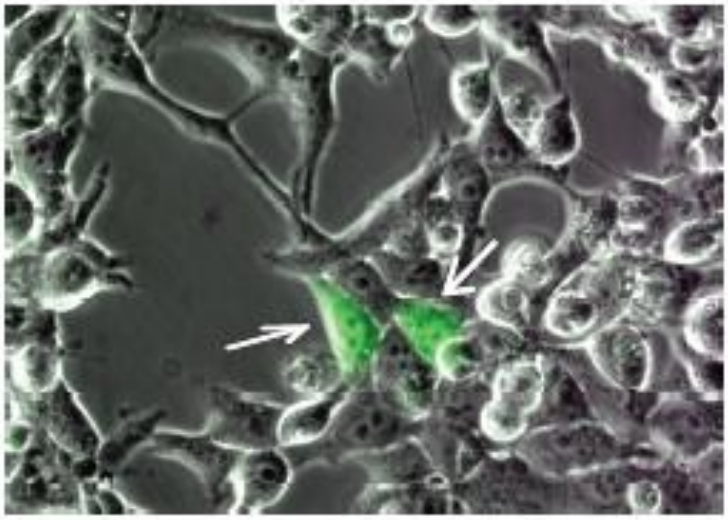

Figure 2: Fluorescent microscope image of a HEK 658D cell after successful oligonucleotide-mediated repair

Figure 2 shows 658D cells fluorescing green under a fluorescent microscope. These cells were treated with the GR66_6R.F repair oligonucleotide during a double-stranded break event, and the apparent presence of a functional GFP gene shows that the repair might have worked as hypothesized.

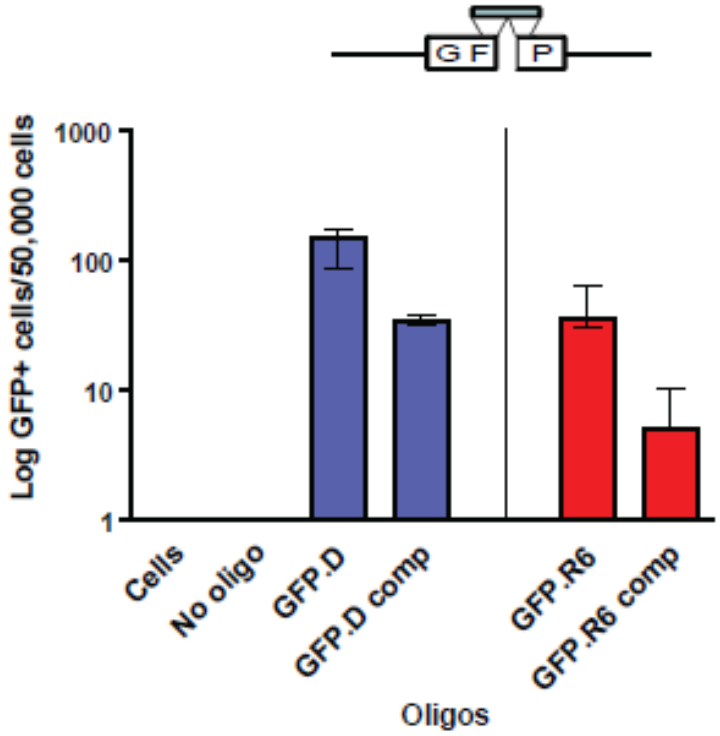

Figure 3: Repair frequencies for cells exposed to a variety of oligonucleotides during an induced DSB.

Figure 3 shows the repair frequencies of each oligonucleotide, based on the number of GFP+ cells in a sample after transfection. The all-DNA and RNA-containing oligonucleotides appear to have similar strand bias for the forward over the reverse-complimentary strands.

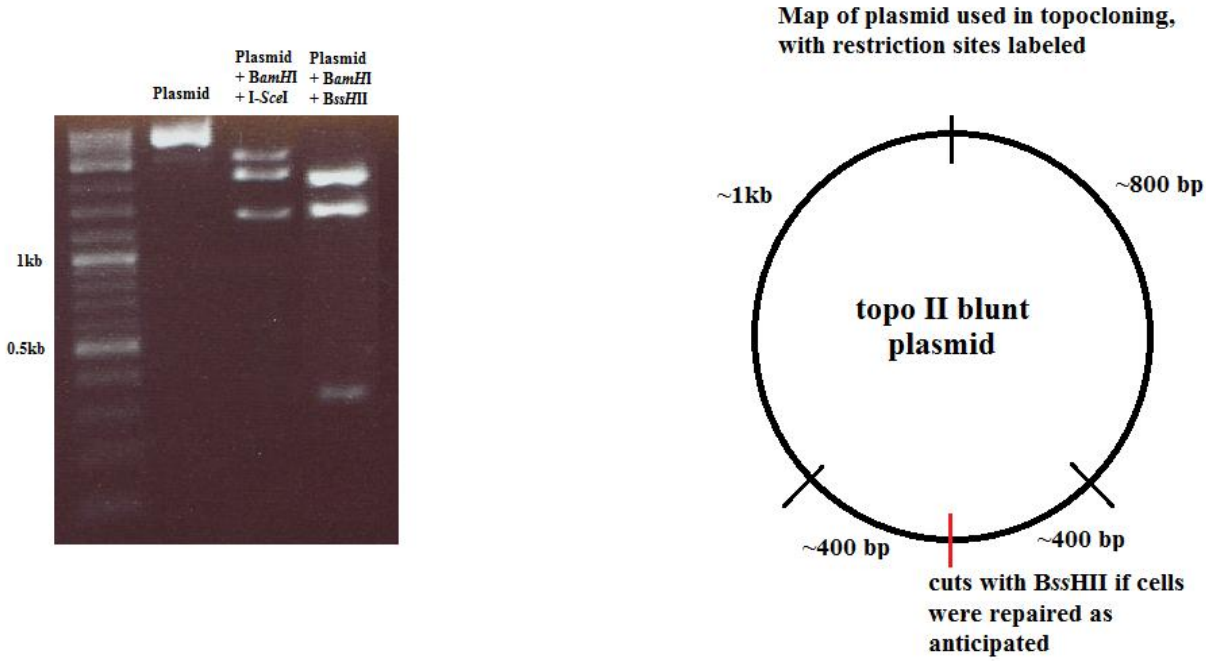

Figure 4: Digestion of topoclone colony plasmid extraction to confirm incorporation of single-nucleotide change during repair.

Figure 4 shows digestion of plasmid extracted from topoclone colony #17 to confirm repair with modification. Colonies from topocloning reaction were inoculated in LB+Kanamycin media and grown for 8 hours. Midi-prep kit (quiagen) protocol was used to extract plasmid from each of these inoculations. Extracted plasmid was digested with either B*amH*I + I-*Sce*I or B*amH*I + B*ssH*II and run through a 1.5% agarose gel to observe digestion results. If repair did not take place as expected, then plasmid extract should show a band ~350bp area upon digestion with I-*Sce*I, and if repair did occur as expected there will be a similar band with digestion of B*ss*HII. The gel show that template repair with incorporation of the changed oligonucleotide did occur.

Chapter 5

Discussion

Figure 3 provides evidence that RNA-templated DNA repair is indeed mediated by an RNA molecule and not a cDNA intermediate by showing that the strand bias of repair is the same for RNA and DNA. In further research, it would be interesting to examine which enzymes are involved in RNA-templated DNA repair using RNAi. By knocking down genes for proteins known to aid in other templated repair pathways (such as homologous recombination) and examining the effects this has on repair frequencies, one could determine the most important proteins in this process in mammalian cells.

This information can hopefully further our understanding of DNA-RNA interactions. It has been shown that ribonucleotides are incorporated into genomic DNA during replication (McElHinny SA et al. 2010). Researchers are inquiring into how this effects the stability of genomic DNA, how enzymes are able to interact with ribo-deoxyribonucleotide chimeras, or more. Future research could examine whether areas with higher-than-average ribonucleotide incorporation have higher mutation rates during replication, lower effectiveness of homologous repair pathways, or other differences from regions with lower incorporation of ribonucleotides.

It was long thought that RNA acted only as an intermediate between DNA and the synthesis of proteins. However, it has been shown that living organisms produce an abundance of non-protein-coding RNA molecules transcribed from “non-coding regions”, which perform a wide variety of functions in the cell. Researchers are discovering new actions these RNA can perform, including RNAs that act as transcription regulators, enzymes, and more (Mercer et al. 2011).

Could there be non-coding RNAs that act in repair pathways as templates? Perhaps there are non-coding RNAs that do not directly act as a template, but are in another way capable of interacting with DNA-associated enzymes. The field of studying non-coding RNAs is still in its infancy, and broadening our knowledge of RNA's capability to interact with DNA molecules is vital to our moving forward in fully understanding how life functions on a molecular level.

Works Cited

Caplen, N. J. and S. Mousses (2003). Short interfering RNA (siRNA)-mediated RNA interference (RNAi) in human cells. Therapeutic Oligonucleotides - Antisense, Rnai, Triple-Helix, Gene Repair, Enhancer Decoys, Cpg and DNA Chips. Y. S. ChoChung, A. M. Gewirtz and C. A. Stein. **1002:** 56-62.

Chayot, R., B. Montagne, et al (2010). "An end-joining repair mechanism in Escherichia coli." Proceedings of the National Academy of Sciences of the United States of America **107**(5): 2141-2146.

Crick, F. (1970). "CENTRAL DOGMA OF MOLECULAR BIOLOGY." Nature **227**(5258): 561-&.

Diciacco, et al. "Site-directed inhibitionof DNA replication by triple helix formation" The Federation of American Societies for Experimental Biology journal **107**(5): 2141-2146.

Fire, A. Z. (2007). "Gene silencing by double-stranded RNA (Nobel lecture)." Angewandte Chemie-International Edition **15**: 2660-2668.

Hershey, A. D. and M. Chase (1952). "INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE." Journal of General Physiology **36**(1): 39-56.

Karagiannis, T. C. and A. El-Osta (2004). "Mechanism of RNA interference, in vivo and potential clinical applications." Cancer Biology & Therapy **3**(11): 1069-1074.

Ren, J. S., X. G. Qu, et al. (2001). "Molecular recognition of a RNA : DNA hybrid structure." Journal of the American Chemical Society **123**(27): 6742-6743.

Shen, Y. et. al. (2010). "RNA-driven genetic changes in bacteria and in human cells." Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis (approved, awaiting publication)

Storici, F., K. Bebenek, et al. (2007). "RNA-templated DNA repair." Nature **447**(7142): 338-341.

Storici, F., J. R. Snipe, et al. (2006). "Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing." Molecular and Cellular Biology **26**(20): 7645-7657.

Mercer, et al. (2009). "Long non-coding RNAs: insights into functions" Nature Reviews Genetics **10**: 155-159

Nick McElhinny SA, et al. (2010). "Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases." PNAS **107**:4949–4954.

Nick McElhinny SA, et al. (2010). "Genome instability due to ribonucleotide incorporation into DNA" Nature Chemical Biology 22 August 2010

Watson, J. D. and F. H. C. Crick (1953). "Molecular structure of nucleic acids - A structure for deoxyribose nucleic acid." Nature **171:** 737.

Zaitsev, E. N. and S. C. Kowalczykowski (2000). "A novel pairing process promoted by Escherichia coli RecA protein: inverse DNA and RNA strand exchange." Genes & Development **14**(6): 740-749.